# 'Super' or just 'above average'? Supershedders and the transmission of *Escherichia coli* O157:H7 among feedlot cattle.

Simon EF Spencer, Thomas E Besser, Rowland Cobbold<sup>‡</sup> and Nigel P French<sup>§</sup>

# Supplementary information

#### Contents

| 1        | MCMC algorithm                       | <b>2</b> |
|----------|--------------------------------------|----------|
| <b>2</b> | Positive tests                       | 4        |
| 3        | Shedding levels                      | 8        |
| 4        | Posterior colonization probabilities | 29       |
| 5        | Parameter posteriors                 | 50       |
| 6        | Extended model: Parameter posteriors | 52       |

\*Dept of Statistics, University of Warwick, UK. Email: s.e.f.spencer@warwick.ac.uk

<sup>†</sup>Dept Veterinary Microbiology and Pathology, Washington State University, Pullman WA 99164 <sup>‡</sup>School of Veterinary Science, University of Queensland, Gatton QLD 4343

 $<sup>\</sup>ensuremath{\S{\text{EpiLab}}}$  Infectious Disease Research Centre, Institute of Veterinary Animal and Biomedical Sciences, Massey University, New Zealand

#### 1 MCMC algorithm

The algorithm of O'Neill and Roberts [1], which was originally intended to update the missing infection times of individuals in a continuous time SIR epidemic model when the removal times are known, requires some minor modifications before it can be applied successfully to our discrete time SIS epidemic model with repeat sampling. In the original algorithm there are 3 steps.

- 1. **Move** A infection time is proposed to be moved to some other time drawn uniformly at random from all permissible times.
- 2. Add An additional infection time is proposed to be added, chosen uniformly at random from all permissible times.
- 3. **Remove** An infection time is proposed to be removed.

In each case the proposal is either accepted or rejected according to the usual Metropolis-Hastings acceptance probability. Importantly the Add and Remove steps are mutual inverses, which allows the probability of the reverse jump being proposed to be calculated. The Move step is a self-inverse.

It is fairly trivial to adapt this procedure from an SIR model with known removal times to an SIS model with unknown removal times, simply by allowing multiple infections and treating the removal times along the same lines as the infection times. In the 'Move' step the removal times are updated in exactly the same way as the infection times. In the 'Add' step both an infection time and removal time are generated together, with the constraint that the removal time must be after the infection time. If the removal time occurs outside of the study period then the individual is assumed to remain infectious. Since the negative binomial distribution for the infectious periods leaves us with a non-Markovian model, and so to calculate the likelihood we need to make an assumption about the time of infection for the individuals that are infected when the study begins. For simplicity, we have assumed that they are all infected on day 1. In the 'Remove' step, both the infection time and removal time (if present) are deleted together.

However, the repeat sampling in our study produces an added difficulty that prevents this algorithm from reaching every possible epidemic state. For example suppose that a particular individual has two positive tests separated by several days during which the individual was not tested. Further, suppose that the Markov chain starts from a state in which the individual is infected throughout this period. It may also have been the case that the individual managed to clear the infection after the first test, but was reinfected before the second test. Since we have assumed that the tests have perfect specificity, any proposal that suggests the individual was not infected on either of the test days will automatically be rejected. It is therefore not possible to construct a series of steps consisting of 'Add', 'Remove' and 'Move' that allows the Markov chain to travel between these two states.

To overcome this limitation we generalised the 'Add' and 'Remove' proposals so that, rather than just adding and removing periods of infection, they could also add and remove periods of susceptibility. More precisely, in the 'Add' step, we choose a period during which a given individual did not change their infection status and then proposed that for a subset of this period their infection status was reversed. Likewise in the 'Remove' step we select an entire episode during which their infection status was unchanged and reverse this whole period, therefore joining the two neighbouring periods together.

The efficiency of this algorithm depends on the size of the blocks that are proposed to be reversed. If the blocks are too large then they no longer relate to the data and so are almost always rejected; whilst if the blocks are too small the proposals ignore the autocorrelation structure present in the individual's colonization history. In order to control and optimize the block size, we imposed a maximum number of days that could be changed by a single proposal, denoted by M. By examining pilot runs with M between 1 and 20, we found that M = 4produced the largest average number of days accepted for reversal. This was approximately the same as the average number of days between tests.

The precise details of our algorithm are given below. Let  $\mathbf{t} = (t_1, t_2, \dots, t_C)$  be the times at which the colonization status of a given animal changes, i.e. for each  $k \in \{1, \dots, C\}$ ,  $X_{p,i,t_k} \neq X_{p,i,t_k+1}$ .

1. Move

- (a) Choose k uniformly at random from  $\{1, \ldots, C\}$ .
- (b) Let  $A = (\{t_{k-1}+1, \ldots, t_{k+1}-1\} \cap \{t_k-M, \ldots, t_k+M\}) \setminus \{t_k\}$  be the set of acceptable times to propose.
- (c) Choose  $t'_k$  uniformly at random from A, and form  $t' = (t_1, \ldots, t'_k, \ldots, t_C)$ .
- (d) Let  $A' = (\{t_{k-1}+1, \ldots, t_{k+1}-1\} \cap \{t'_k M, \ldots, t'_k + M\}) \setminus \{t'_k\}$  be the set of acceptable times to propose from  $t'_k$ .
- (e) Accept proposal  $t_k \mapsto t'_k$  with probability  $\min\{1, \frac{\pi(t')|A|}{\pi(t)|A'|}\}$ , where  $\pi$  is the posterior density.

- 2. Add
  - (a) Let  $B = \{t : X_{p,i,t} = X_{p,i,t+1} = X_{p,i,t+2}\}$  be the set of acceptable beginning times of a new block.
  - (b) Choose b uniformly at random from B.
  - (c) Let  $n = \min\{k : t_k > b\}$  be the next block.
  - (d) Let  $E = \{b + 1, \dots, t_n 1\} \cap \{b + 1, \dots, b + M\}$  be the set of acceptable end times of a new block starting at b.
  - (e) Choose *e* uniformly at random from *E* and form  $t' = (t_1, \ldots, t_{n-1}, b, e, t_n, \ldots, t_C)$ .
  - (f) Let R' be the set of acceptable blocks that can be removed from t' (see Remove step).
  - (g) Accept proposal  $\boldsymbol{t} \mapsto \boldsymbol{t}'$  with probability  $\min\{1, \frac{\pi(\boldsymbol{t}')|B||E|}{\pi(\boldsymbol{t})|R'|}\}$ .
- 3. Remove
  - (a) Let  $R = \{k : t_k + M \ge t_{k+1}\}$  be the set of acceptable blocks that can be removed from t.
  - (b) Choose r uniformly from R and form  $\mathbf{t}' = (t_1, \ldots, t_{r-1}, t_{r+2}, \ldots, t_C)$ .
  - (c) Let B' be the set of acceptable beginning times of a new block in t' (see Add step).
  - (d) Let E' be the set of acceptable end times of a new block in t' starting at  $t_r$  (see Add step).
  - (e) Accept proposal  $\boldsymbol{t} \mapsto \boldsymbol{t}'$  with probability  $\min\{1, \frac{\pi(\boldsymbol{t}')|R|}{\pi(\boldsymbol{t})|B'||E'|}\}$ .

For the purposes of the Add and Remove steps in this algorithm  $t_1 = 0$  and  $t_C = T$  (the final day in the study) to allow incomplete blocks to be added and removed at the ends of the study period.

#### 2 Positive tests

*E. coli* O157:H7 positive tests for each of the 20 study pens. The letters "SS" indicate that the animal was identified as being a supershedder in [2].







# 3 Shedding levels

Measured shedding levels of *E. coli* O157:H7 and exponentially smoothed shedding levels for each animal in the study. Plots are arranged so that each page constitutes the animals from a single pen. The letters "SS" indicate that the animal was identified as being a supershedder in [2].















Shedding levels, Pen 15 (North), Animal 2



















![](_page_23_Figure_0.jpeg)

![](_page_24_Figure_0.jpeg)

![](_page_25_Figure_0.jpeg)

![](_page_26_Figure_0.jpeg)

![](_page_27_Figure_0.jpeg)

# 4 Posterior colonization probabilities

Posterior colonization probability with positive and negative tests marked at 1 and 0 respectively for the model described in Main Text Section 2.5. In this model the value of the supershedding threshold  $\tau$  was inferred from the data. Plots are arranged so that each page constitutes the animals from a single pen. The letters "SS" indicate that the animal was identified as being a supershedder in [2].

![](_page_29_Figure_0.jpeg)

![](_page_30_Figure_0.jpeg)

![](_page_31_Figure_0.jpeg)

![](_page_32_Figure_0.jpeg)

![](_page_33_Figure_0.jpeg)

40 60

Time (days)

![](_page_34_Figure_0.jpeg)

![](_page_35_Figure_0.jpeg)

Time (days)

![](_page_36_Figure_0.jpeg)

![](_page_37_Figure_0.jpeg)

![](_page_38_Figure_0.jpeg)

![](_page_39_Figure_0.jpeg)

![](_page_39_Figure_1.jpeg)

![](_page_39_Figure_2.jpeg)

![](_page_40_Figure_0.jpeg)

![](_page_41_Figure_0.jpeg)

![](_page_42_Figure_0.jpeg)

![](_page_43_Figure_0.jpeg)

![](_page_44_Figure_0.jpeg)

![](_page_45_Figure_0.jpeg)

![](_page_45_Figure_1.jpeg)

![](_page_46_Figure_0.jpeg)

![](_page_47_Figure_0.jpeg)

![](_page_48_Figure_0.jpeg)

Time (days)

60 40

### 5 Parameter posteriors

Posterior distributions for the parameters of the model described in Main Text Section 2.5. In this model the value of the supershedding threshold  $\tau$  was inferred from the data. The final two plots show the joint posterior distribution of the supershedding threshold  $\tau$  and the relative risk from supershedders  $\rho$  – the first shows the posterior samples whilst the second shows a 2-dimensional Kernel Density Estimate (KDE).

![](_page_49_Figure_2.jpeg)

Posterior distribution of the mean colonization duration

Posterior of the shape parameter of colonization period

![](_page_50_Figure_2.jpeg)

Posterior distribution of the sensitivity of RAMS

![](_page_50_Figure_4.jpeg)

![](_page_50_Figure_5.jpeg)

Posterior distribution of the sensitivity of faecal test

![](_page_50_Figure_7.jpeg)

![](_page_50_Figure_8.jpeg)

Relative risk from supershedders

![](_page_50_Figure_9.jpeg)

0.02

0.04

5.0

0.08

![](_page_50_Figure_10.jpeg)

Density

#### 6 Extended model: Parameter posteriors

Posterior distributions for the parameters of the the model described in Main Text Section 2.6. In this model faecal material accumulates in the environment once it has been shed, before decaying away. This allows individuals to become colonized from animals that were shedding earlier in the study but are not shedding contemporaneously. The value of the supershedding threshold  $\tau$  was inferred from the data. The final plot is a rescaled version of the posterior distribution fo the risk decay parameter  $\delta$ .

![](_page_51_Figure_2.jpeg)

![](_page_52_Figure_0.jpeg)

Posterior distribution of the sensitivity of RAMS

![](_page_52_Figure_2.jpeg)

![](_page_52_Figure_3.jpeg)

![](_page_52_Figure_5.jpeg)

![](_page_52_Figure_6.jpeg)

Density

Posterior distribution of the sensitivity of faecal test

![](_page_52_Figure_8.jpeg)

Posterior distribution of risk half-life

![](_page_52_Figure_10.jpeg)

## References

- O'Neill PD, Roberts GO. 1999 Bayesian inference for partially observed stochastic epidemics. J. R. Stat. Soc. A 162, 121–129.
- [2] Cobbold RN, Hancock DD, Rice DH, Berg J, Stilborn R, Hovde CJ, Besser TE. 2007 Rectoanal junction colonization of feedlot cattle by *Escherichia coli* O157:H7 and its association with supershedders and excretion dynamics. *Appl. Environ. Microbiol.* 73, 1563–1568.